ChinaXiv合作期刊

chinaXiv:202108.00051v1

使用 Biotracks 采集植物标本

徐洲锋[1,2]，刘恩德[2]，陈家辉[2*]

（1. 西南林业大学 林学院，昆明 650224；2. 中国科学院昆明植物研究所 东亚植物多样性与生物地理学重点实验室，昆明 650201）

摘要：Biotracks 是一款自然观察类的公众科学应用，目前已经被各类科学调查和自然观察项目广泛使用。我们利用 Biotracks 的标本采集项目将野外采集的数据与标本馆的数字馆藏系统连接起来，使用户在手机上记录的信息，可以被应用到标本馆的标本数字化之中。这种方式不仅提升了数字标本的转录效率，还从根本上改变了整个标本收集流程中的数据整合方式，使得标本从采集到收藏的各个环节都能够获得高质量的效率提升。同时，新的标本收集模式还能够自然地将标本的野外照片与数字标本融为一体，从而使得传统标本原本很难呈现的颜色、行为、立体结构、环境等信息最终可以通过数字标本再次展现给研究者。这不仅在信息维度上拓展了传统标本的内涵，结合公众科学，未来还有望进一步延伸馆藏标本鉴定和讨论的时空范围。除此之外，公众科学在解决标本馆问题中所展现出来的潜质，也为我们重新审视标本馆的领域价值提供了新的视角。

关键词：Biotracks，标本馆，植物标本，标本数字化，公众科学

中图分类号： **文献标识码：**A

Using the Biotracks platform to collect plant specimens

XU Zhoufeng[1,2], LIU Ende[2], CHEN Jiahui[2*]

（1. *College of Forestry*, *Southwest Forestry University*, Kunming 650224, China; 2. *Key Laboratory for Plant Diversity and Biogeography of East Asia*, *Kunming Institute of Botany*, *Chinese Academy of Sciences*, Kunming 650201, China）

Abstract: Biotracks is a citizen science platform of nature observation, which has been widely used in the scientific field survey and public nature observation in China. We use the collection object of Biotracks to link the data collected in the field to the digital collection system of the herbarium. Users can quickly record the information on mobile phone in the wild and then apply these data to specimen digitization. This new method improves the efficiency and accuracy of all steps: from the collecting event to specimen curation. More importantly, this method can link field photos with specimen data naturally, which means that the color, behavior, three-dimensional structure, environment and other information that the traditional specimens are difficult to display

---

基金项目：云南省重大科技专项计划-电子信息与新一代信息技术重大专项（2018ZI001-3）；中国科学院战略性先导科技专项（A 类）地球大数据地球大数据科学工程（XDA19050202）；中国科学院战略生物资源计划（KFJ-BRP-017-13）；云南省基础研究专项（202101AS070032）[Supported by Yunnan Key Science And Technology Special Projects-New Generation of Information Technology (2018ZI001-3); Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19050202); Biological Resources Programme, Chinese Academy of Sciences (KFJ-BRP-017-13); Yunnan Fundamental Research Projects (202101AS070032)]。

作者简介：徐洲锋(1989 -)，硕士，助理工程师，研究方向为生物多样性信息学和生物地理学，（E-mail）xuzhoufeng@mail.kib.ac.cn。

***通信作者：**陈家辉，博士，研究员，研究方向为植物系统演化及生物多样性信息学，（E-mail）chenjh@mail.kib.ac.cn。

ChinaXiv合作期刊

now can be presented to the researchers. This not only expands the value of traditional specimens, but also hopes to extend the time and space of specimen identification and discussion in the future. In addition, the potential of citizen science in solving the problem of herbarium also provides a new perspective for us to redefine the scientific value of herbarium.

Key words: Biotracks, herbarium, specimens, specimen digitization, citizen science

标本数字化支持了诸如 GBIF（http://www.gbif.org/），NSII（http://www.nsii.org.cn/），iDigBio（https://www.idigbio.org/）等全球和国家尺度的生物多样性信息平台以及各类标本馆和区域尺度的标本数据库建设。这些富集的数据结合涌现的新技术（尤其是信息技术和测序技术）进一步促进了各类新应用、新研究的快速发展（Soltis, 2017; Morrison et al., 2017; Willis et al., 2017; James et al., 2018），并为我们在时空尺度上探究生物之间的宏观联系及其微观机理带来了全新的视角和模式。同时，标本的数字化也促进和支持了生物多样性信息学这门新兴学科的诞生和发展（王利松等，2010）并使其在全球、地区和国家尺度的标本数据整合、共享与利用等方面发挥出了中流砥柱的作用。

然而，相比迅速发展的标本数据整合与分享，标本馆收集标本的方式和流程却鲜有变化（Schindel & Cook, 2018）。虽然目前也有一些现代的技术和方法被采用，比如使用 GPS 记录地理坐标与海拔、使用条形码标识标本、采用数据库管理标本，将分子材料纳入采集保藏职能等，但是标本收集流程中固有的缺陷却仍然未能得到有效地解决（Daru et al., 2017; 徐洲锋，2018），这包括：（1）许多标本的信息不完整，导致标本价值受限（比如采集人、采集时间、采集地点等关键信息的缺失）；（2）不同来源的标本，数据差异较大，致使数据整合困难；（3）一些宏观信息难以在标本中得到准确保存（比如花的颜色、分枝的角度等）；（4）各种原因造成的数据偏差（比如错误的日期、坐标值和拉丁名拼写）；（5）重复且低效的标本数据转录；（6）缓慢的标本鉴定与讨论。针对这些缺陷，也有一些解决方案，比如尽可能地标准化传统标本收集流程的各个环节；开发定制化的标本数字化系统（Gries et al., 2014）；采用新设备和新技术生产数字标本（Blagoderov, 2012; Drinkwater et al., 2014）。这些工作确实可以从各个环节提升标本收集的效率和品质，但其本质上仍然属于对传统标本收集流程的优化，因此很难从体系上彻底解决上述问题。

当然，这并不意味着原有的标本收集流程是错误的。事实上它至今仍然能够很好地满足传统分类学的需要。只是由于学科的发展，新的研究方法和研究内容对标本资源，特别是对标本的数字化资源提出了新的要求，比如需要更精准的位置信息、能够很好地执行数据聚合、需要稳定持久的标识、便于分享与利用等（Vissers et al., 2017; Hobern et al., 2019），而原有的模式却很难适应这些新的需求。通过公众科学解决这些问题是目前国际研究的趋势，比如让博物馆（标本馆）支持业余科学家的成长和发展，招募志愿者参与博物馆（标本馆）的野外调查，通过众包转录和审核数字标本（Sforzi et al., 2018）。这些尝试很大程度上展示出了公众科学在博物馆生物多样性研究中的巨大潜质。然而这些工作之间尚未构成体系，因此还不能从整体上重塑标本的收集流程。对此，本文基于 Biotracks 公众科学平台构建并实践了一种新型的标本收集模式，该模式可以提高植物标本收集的效率和品质，并有望系统地解决传统标本馆收集流程中的诸多缺陷。同时结合 Biotracks，本文还简要阐述了根植于公众发展标本馆的必要性，以及标本馆应用公众科学的一些思路，以期为国内同行抛砖引玉。

1 材料与方法

1.1 Biotracks 客户端简介

Biotracks 是由中国科学院昆明植物研究所开发的一款开放免费的自然观察类公众科学应用，该应用可以帮助用户组织自然观察项目，记录户外生物物种信息。Biotracks 目前拥

chinaXiv:202108.00051v1

chinaXiv:202108.00051v1

有 WEB 端（http://www.biotracks.cn）、Android 版和 IOS 版（下载地址：http://www.biotracks.cn/app/biotracks，也可在各类国产手机应用市场自行搜索中文名“生命观察”下载安装），同时微信小程序版也将于 2021 年 6 月份正式上线。本研究中，我们使用 Biotracks 手机客户端记录标本的采集信息，使用 Biotracks WEB 端进行数据的管理和下载。为了保证 APP 能够彻底地替代纸媒或电子表格记录，Biotracks 手机客户端专门为植物标本的采集记录做了以下一些适配和优化。

1.1.1 定制专用记录模版 Biotracks 能够为不同类型的任务定制专属的数据记录模版（比如鸟类观察、植被样方/样线调查等），我们根据标本采集记录的需要，专门拟定了相应的数据标准（详见下一节），并以此制定了 Biotracks 的标本采集记录模版。

1.1.2 自动记录关键信息 APP 能够自动记录观察对象的地理坐标、海拔、发现时间、记录编号等信息，且这一过程并不需要网络的支持。同时，APP 在有网络状态下还可以利用地理坐标解析获得结构化的地址信息。

1.1.3 快速录入物种学名 通过网络，APP 用户可以使用中文俗名、俗名拼音、拉丁名等字符检索到逾百万条的全球植物、菌物、鸟类、鱼类以及中国两栖、爬行和哺乳动物的拉丁名以供使用；而在无网络环境下，用户也可以通过 APP 内置的十万多条中国生物物种名称库检索所需要的拉丁名，以满足国内大多数离线场景下的学名录入需求。

1.1.4 支持 AI 辅助鉴定 APP 的人工智能图像识别引擎会尝试识别所拍摄的物种照片，并推荐可能的拉丁名以供用户参考使用。

1.1.5 高效的人机交互 为了保证手机操作的便捷性，Biotracks APP 的记录界面做了大量的人机交互优化，这包括：(1)界面中每个字段所处的位置都会被仔细考量，以尽可能降低信息录入时手指划动的范围（比如通常情况下 APP 会自动写入海拔、地理坐标等信息，这些无需人工录入的字段就会被放置在录入界面的后部，以减少干扰）；(2)不同的字段会选择不同的录入方式，不同字段的内容录入还会辅以不同的快捷键，以尽可能简化录入操作（比如有些字段会采用下拉选项写入文本，有些字段会直接调用数字键盘键入数字，而有些字段则可以直接通过“复制”快捷键复制上一条记录的内容）；(3)对于一些内容比较稳定的字段，APP 会自动复用上一条记录的内容，以减少人工操作（比如同一次采集活动中，采集人员的姓名大部分情况下是不会改变的，记录模版会自动将上一条记录的采集人员信息写入下一条记录）。

1.1.6 数据安全的保障 为了防止因用户手机丢失、损坏、更换导致的数据遗失。手机 APP 记录的所有数据都可以在有网络的情况下上传至 Biotracks 云端。用户可以通过个人账户登陆 biotracks.cn 访问、编辑、下载这些数据。而如果用户更换了手机，云端的数据也会加载到新手机的 APP 之中。

ChinaXiv合作期刊





图 1 Biotracks 官方网站 biotracks.cn 首页

chinaXiv:202108.00051v1

ChinaXiv合作期刊

Fig.1 The homepage of biotracks.cn

1.2 数据标准

Biotracks 可以根据不同的任务需求定制相应的数据模板。对于标本采集，我们遵循 DarwinCore（http://rs.tdwg.org/dwc/）定义了 14 个核心字段（表 1）。显然这 14 个字段并不能覆盖完整的标本记录信息（比如缺少行政区划、鉴定人、鉴定时间、科、属等等），但其他必要的信息一般都可以根据这些字段的值以及用户的 ID 由 Biotracks 自动推断生成和补充。这样的设计可以让用户录入最少的内容，实现最大的信息完整性。同时，基于这些核心字段，我们还为不同分类群分别定义了一套性状描述属性，这些属性，会被定义为一种形似字段的数据库虚拟列(https://dev.mysql.com/doc/refman/5.7/en/json.html)，并以 JSON 对象的形式存储在 dynamicProperties 核心字段下。对于用户而言这些属性与其他字段无异，但却能极大地简化和统一不同类群的数据模版设计，并提高应用适配的灵活性。

借助以上方案，Biotracks 不仅大幅减少了人工记录的内容、降低了用户界面 (User Interface, UI) 的复杂度，还统一了各个类群的数据记录标准，从而为标本数据的跨类群、跨平台整合提供了良好的基础。本研究主要针对植物标本的采集，因此使用植物相关的模版进行数据的记录。

表 1 Biotracks 标本记录项目定义的 14 个核心字段

Table 1 14 core fields in specimens collection project of Biotracks

字段名称 Field name	中文名对照 Chinese name	示例 Example
category	大类别	Pteridophyta
recordedBy	采集人	刘恩德，徐洲锋
recordNumber	采集号	9170
eventDate	采集时间	2018-12-12 13:33:04
locality	小地点	松林沟
decimalLatitude	纬度	46.9833486
decimalLongitude	经度	85.9428474
minimumElevationInMeters	海拔（低值）	1 816
maximumElevationInMeters	海拔（高值）	1 840
habitat	生活环境	多石山坡灌丛
occurrenceRemarks	采集备注	当地药用
dynamicProperties	动态属性	JSON 对象，属性可依据类群不同而不同，比如对于种子植物，它包含了：根、茎、叶、花、果、种子、胸径、体高、频度等扩展属性。
associatedMedia	关联媒体	http://cloudfile.biotracks.cn/userimg/28/2019-08-16/74498871565930525.jpg!740
scientificName	学名	*Cystopteris fragilis*

注：除了 category 是根据需要自行扩展获得，其他字段名均来源于 Darwin Core 标准。

Note: Most fields are named according to the Darwin Core Standard, except for ‘category’ is obtained according to actual needs.

1.3 数据采集

chinaXiv:202108.00051v1

chinaXiv:202108.00051v1

使用 Biotracks 执行标本采集信息的记录，需要首先在手机 APP 上建立一个标本采集项目，这个项目可以是公开的（公共项目，所有用户可见可加入）、部分可见的（团队项目，只有创建者才能邀请人员加入）或者私密的（私人项目， 只有创建者可以使用）。加入项目的成员会使用一致的数据模版记录标本采集信息。采集标本时，记录者需要首先预设第一条记录的采集人、采集号样式及其起始序号等通用信息。后续记录会自动写入上一条记录中需要复用的信息，并能依据预设的编号按序编排采集号。当用户在标本采集的现场记录信息时，APP 会默认写入当前的海拔、经纬度和时间信息，用户也可以手动编辑这些信息，或者通过所拍摄的现场照片重置这些信息（若照片有 GPS 坐标，用户可以使用照片的 GPS 坐标覆写当前坐标数值，这对于后补采集记录非常有用）。记录时，手机拍摄的照片会与采集信息直接绑定，同时 AI 引擎还会识别所拍摄的照片，并给出物种的鉴定参考，用户可以直接使用或修改机器推荐的鉴定。(使用 Biotracks APP 记录标本采集信息的详细说明请见：https://mp.weixin.qq.com/s/daQRSbAA9sRX1T5FNfIOkQ）。

1.4 数据的编辑与整合

标本采集活动通常会由多人或多个队伍组成，其中每个成员的工作必然会各有侧重。比如一些成员擅长鉴定，另外一些负责摄影，还有一些需要负责材料的处理与标记。在传统模式下，这种零散而又各自相关的工作很难实现跨时空的作业配合，整合这些工作产生的数据更是相当困难。通过 Biotracks 的项目协同功能则可有效地解决决这一问题。比如实际的标本采集活动中，采集信息的整理通常会是一个相对滞后的工作，而 Biotracks 则可以将项目内每个成员记录的采集信息通过网络汇总到云端，以供所有成员在电脑和手机端浏览和编辑。同时，一些成员作出的鉴定，也可以通过云端传递给其他成员参考或使用，而每个成员对于数据的编辑和补充，同样也会及时反馈到云端。这种通过信息共享连接线下与线上作业的协同模式，能够将项目数据的最新状态充分整合到云端并同步到手机 APP 内，从而让原本独立、分散的个体工作，联结成可以协作互助的动态网络。

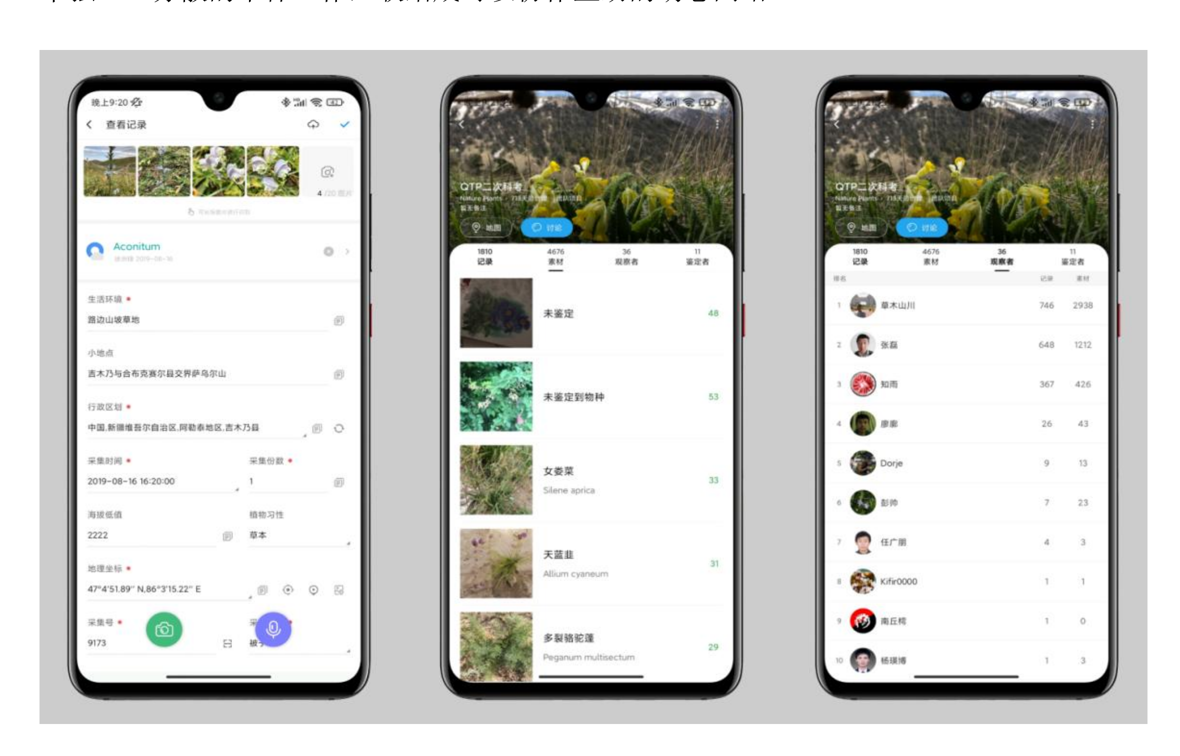


图 2 Biotracks 植物标本采集信息记录界面（左）、项目协同内的图片汇总归类界面（中）与成员数据贡献排名界面（右）

Fig.2 Plant collection information record interface (left), image classification and summary

interface (middle) and member data contribution ranking interface (right) in project collaboration of Biotracks

1.5 连接馆藏数字标本

Biotracks 记录的采集信息可以被导出为符合 Darwin Core（http://rs.tdwg.org/dwc/）规范的 Excel 数据表。这些 Excel 数据表既可以被用于纸质标签的打印，也可以被导入到标本馆的数据库中作为标本数字化的基础资料，以避免标签信息被二次转录。此外，数据表中的 associateMedia 字段还记录了标本的野外照片地址，标本数字化系统可以借此实现数字标本和标本原生照片的自然整合。

标本馆的数字化系统需要依据 Biotracks 采集信息的数据规范做一些技术适配，才可以导入 Biotracks 的数据表。考虑到目前国内标本馆信息化建设的现状，我们专门开发了一套专业的数字标本馆内容管理系统（Content Management System, CMS）：Kingdonia 数字标本馆系统（徐洲锋，2018）。该系统高度适配了 Biotracks 的数据标准和接口协议，不仅可以导入 Biotracks 数据表，还可以通过 Biotracks 授权的接口检索和获取相应的采集记录（图 3），这样即便标本馆在接收标本时没有获得相应的电子数据，也可以调用存储在 Biotracks 中的采集信息以提升标本数字化的效率。此外，Kingdonia 系统还支持独立部署和管理，系统能够在线获得升级，并可以根据不同标本馆的馆藏类群和作业模式进行定制开发，因此可以被用于对数据的权属以及管理都有较高要求的标本馆（博物馆）。

chinaXiv:202108.00051v1

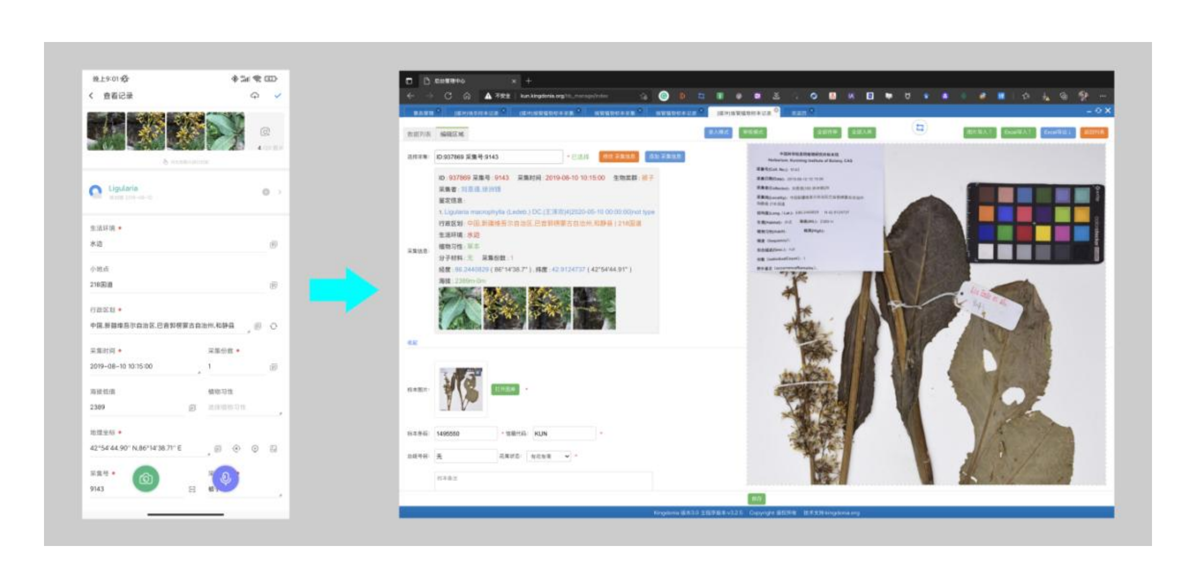


图 3　Kingdonia 可以调用 Biotracks 记录的采集信息

Fig.3 Kingdonia can call the collection information recorded by Biotracks

1.6 效率评估

Kingdonia 系统支持多种标本数字化模式，工人既可以按照传统方式对照标本标签转录数字标本；也可以从 Biotracks 调取数据以用于数字标本的转录。为了评估这两种模式的转录效率差异，我们从 Kingdonia 系统中导出了昆明植物研究所标本馆（KUN）两位工人的标本转录日志，该日志记录了每一份标本开始转录和结束转录的时间，因此可以被用于标本转录效率的量化分析。

不同标本的转录效率是不一致的，为了减少标本因素的影响，我们尽可能提取两位工人相邻时间的转录日志，以确保不同工人转录的标本，其性质尽可能接近。每位工人的日志又会被分为人工转录的标本（I 类样本）和调用 Biotracks 数据转录的标本（II 类样本），最终根据人员不同总共可获得 A-I、A-II、B-II、B-II 四组样本（A、B 指代不同的工人）。每组样本都可以根据标本的开始转录时间和结束转录时间计算出每份标本的转录时长，由于这

chinaXiv:202108.00051v1

些标本都是在日常工作中被自然转录，因此会存在一些转录时长明显脱离实际的标本（比如工人在转录的中途开始休息，或者临时有事打断了原有的转录工作都可以造成标本转录时间被延长）。根据经验，我们将各组中转录耗时最长的前 10% 的标本作为无效值排除，剩下的标本作为可评估样本参与量化分析。

对于可评估样本，需要计算出每组样本中标本转录时长的中位数、上下四分位数、上限、下限、平均值，其中上限采用上四分位数加 1.5 倍四分位距获得，下限采用下四分位数减 1.5 倍四分位距获得。然后对每组样本做箱体图进行比较分析。

2 结果与分析

2.1 Biotracks 在植物标本采集活动中的推广效果

Biotracks 的构想始于 2013 年，并于 2014 年底与 Kingdonia 系统同步开始实施。整个平台于 2016 年 9 月对外正式上线了首个 WEB 端（http://www.kib.cas.cn/xwzx/zhxw/201609/t20160907_4659604.html），于 2017 年 9 月发布了首个手机版本（http://www.kib.cas.cn/xwzx/zhxw/201709/t20170901_4854058.html）。后经若干年的发展，目前 Biotracks 共有注册用户 16 987 名，这些用户总共记录了 200 多万张的物种照片、60 万条的物种观察记录，其中近 23 万条为植物标本的采集记录，有关 Biotracks 的用户、物种观察记录、图片、植物标本采集记录的增长趋势请见图 4。

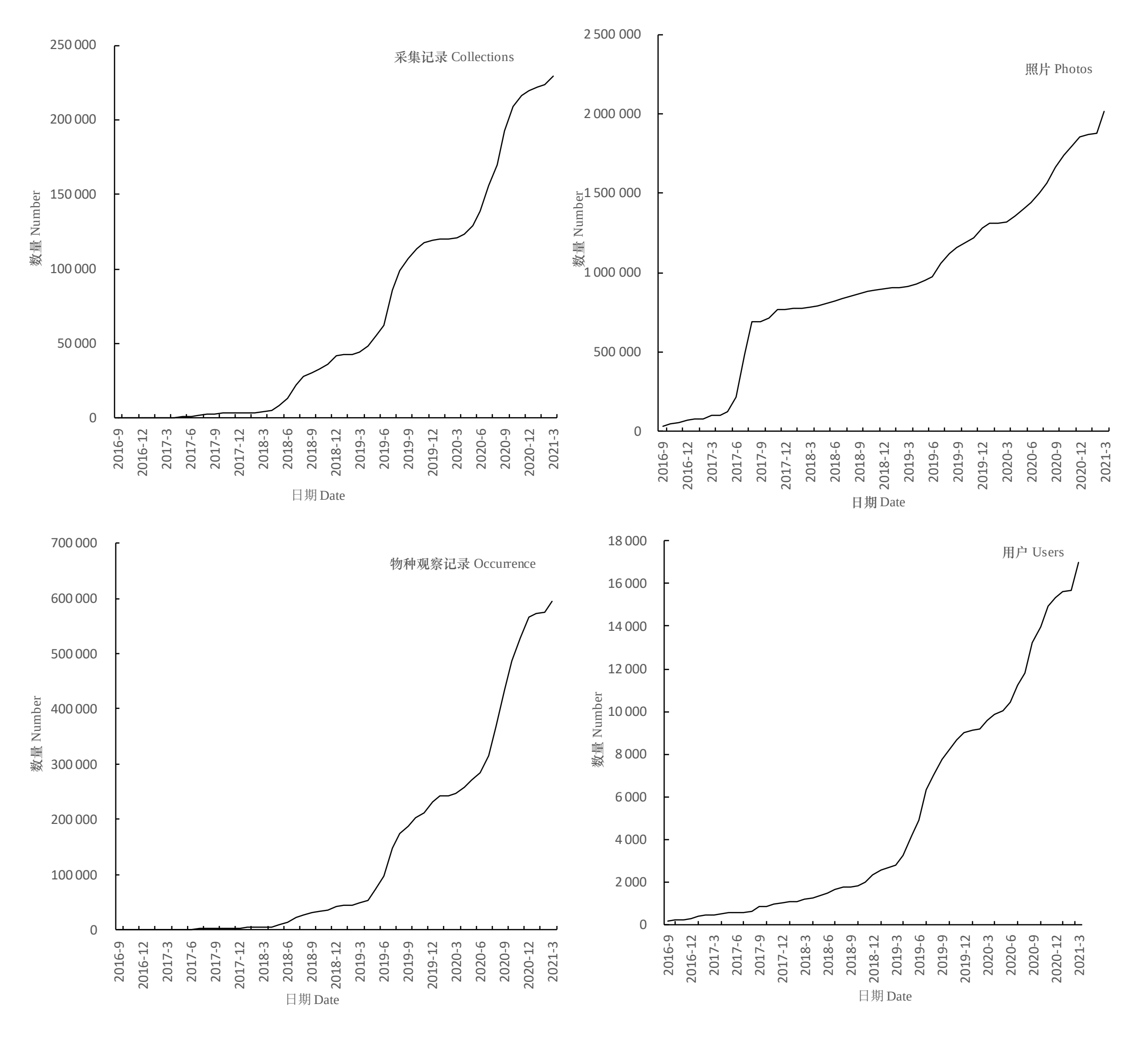


图 4　2016 年 9 月至 2021 年 3 月，Biotracks 标本采集记录（左上）、图片数量（右上）、注册用户（右下）、物种观察记录（左下）增长折线图

Fig.4 Growth line chart of collection record (upper left), photos number (upper right), users (lower right), occurrence record (lower left) from September 2016 to March 2021 on Biotracks

对 Biotracks 的用户群体进行分析，可以发现与标本馆直接相关的用户中，比较有代表性的群体主要来自于中国科学院。其中表 2 罗列出了目前用户最多的前 15 个中国科学院下属机构，可以看出主要是一些与动植物相关的研究所和植物园。这些机构内的植物标本馆不仅各具区域特色，其馆藏标本的总量甚至占据了中国过半的植物标本储量（贺鹏等，2021）。Biotracks 目前也被这些单位中的昆明植物研究所标本馆（KUN）、华南植物园标本馆（IBSC）、西双版纳热带植物园标本馆（HITBC）、成都生物研究所标本馆（CDBI）、新疆生态与地理研究所标本馆（XJBI）、南京中山植物园标本馆（NAS）等标本馆所采用，他们不仅将 Biotracks 应用到了日常的标本采集活动中，还将 Kingdonia 数字标本馆系统引入到馆藏标本的管理之中，从而形成了完整的应用体系（http://www.kib.ac.cn/xwzx/zhxw/201912/t20191231_5479319.html）。

表 2 中国科学院下属单位与中国高等院校中用户数量最多的 15 个机构

Table 2 15 CAS institutions and universities with the largest number of users

机构 Institution	用户数量 No. of users	机构 Institution	用户数量 No. of users
中国科学院昆明植物研究所 Kunming Institute of Botany, CAS	307	浙江大学 Zhejiang University	725
中国科学院西双版纳热带植物园 Xishuangbanna Tropical Botanical Garden, CAS	126	武汉大学 Wuhan University	367
中国科学院华南植物园 South China Botanical Garden, CAS	72	赣南师范大学 Gannan Normal University	285
中国科学院植物研究所 Institute of Botany, CAS	66	淮北师范大学 Huaibei Normal University	217
江苏省中国科学院植物研究所 Institute of Botany, Jiangsu Province, CAS	43	福建农林大学 Fujian A & F University	215
中国科学院新疆生态与地理研究所 Xinjiang Institute of Ecology and Geography, CAS	40	安徽农业大学 Anhui Agricultural University	201
中国科学院武汉植物园 Wuhan Botanical Garden, CAS	34	华南农业大学 South China Agricultural University	193
中国科学院上海辰山植物园 Shanghai Chenshan Botanical Garden	34	贵州中医院大学 Guizhou University of Traditional Chinese Medicine	188
中国科学院成都生物研究所 Chendu Institute of Biology, CAS	33	西南大学 Southwest University	170
中国科学院庐山植物园 Lushan Botanical Garden, CAS	21	山西大学 Shanxi University	164

chinaXiv:202108.00051v1

ChinaXiv合作期刊

中国科学院昆明动物研究所 Kunming Institute of Zoology, CAS	20	宁夏大学 Ningxia University	126
中国科学院西北高原生物研究所 Northwest Institute of Plateau, CAS	12	吉首大学 Jishou University	124
中国科学院沈阳应用生态研究所 Institute of Applied Ecology, CAS	12	南京师范大学 Nanjing Normal University	112
中国科学院动物研究所 Institute of Zoology, CAS	11	兰州大学 Lanzhou University	101
中国科学院东北地理与农业生态研究所 Northeast Institute of Geography and Agroecology, CAS	10	西北农林科技大学 Northwest A & F University	100

注：用户所属机构为用户注册账户时自行提交的文本，一些机构会有很多别称，这里提供的仅是根据机构标准称谓聚类获得的数据。

Note: The institution name is the text submitted by users when they register their accounts. Some institutions will have many nicknames. The data provided here is only based on the standard name of the institution.

除此之外，高校则是 Biotracks 目前用户群体中最大的单一来源。Biotracks 有超过 100 多所高校的用户（表 2 列出了用户最多的 15 所高等院校），这些用户既使用 Biotracks 进行野外科考，也会使用 Biotracks 引导学生进行标本的采集和物种的观察记录，Biotracks 的项目统计则可以将每个学生的实际贡献反馈给老师作为评估实习效果的参考。

而在实际应用中，Biotracks 的标本采集项目可涉及生物多样性调查的方方面面，表 3 列出了一些具有代表性的标本采集项目，这些项目涉及自然保护区的本底资源调查、各种区域尺度的生物多样性编目、特定生物类群的调查、植物志编纂、植物种质资源的采集、中药普查等等。此外，第二次青藏高原科学考察 植物多样性的调查还将 Biotracks 作为植物标本采集的推荐应用进行了推广。应该说，Biotracks 在国内植物标本的采集活动中，目前已经被领域广泛接纳和使用。

表 3 Biotracks 上一些具有代表性的标本采集项目

Table 3 Some typical specimen collection projects of Biotracks

项目名称 Project name	采集号数 No. of collection
贵大-贵州省菌物资源普查及创新利用 Guizhou University - General Survey and Innovative Utilization of Fungi Resources in Guizhou Province	9 729
2020 年度地衣二次科考 Scientific Investigation of Lichen - QTP2020	4 520
大别山东片区苔藓调查 A Survey of Bryophytes in Eastern Dabie Mountains	1 041
蕨类采集(汪) Fern Collection by Wang	1 946
青海南-川西高山冰缘带植物科考	3 750

chinaXiv:202108.00051v1

Scientific Investigation of Plants in Alpine Periglacial Zone of Southern Qinghai and Eastern Sichuan	
西藏科考	
Scientific Investigation of Xizang Province	2 767
大别山维管植物	
Vascular Plants in Dabie Mountains	761
梵净山水青冈采集	
Fagus longipetiolata Collection in Fanjin Mountain	57
世界科级植物采集	
Collection of World Family Plants	312
豆-Legume Project	
Fabaceae-Legume Project	2 985
冬青采样	
Ilex Sampling	616
棕榈科条形码研究	
Barcoding Research of Arecaceae	314
HITBC 标本联盟	
Specimen Alliance of HITBC	4 729
种质库采集部	
Collection Department of GBOWS	1 475
西双版纳国家级自然保护区本底调查	
Background Survey of Xishuangbanna National Nature Reserve	3 049
QTP 天山帕米尔科考分队 2020	
Tianshan Pamirs Scientific Expedition - QTP2020	1 344
藏东南植物调查	
Plant Survery in Southeast Xizang Province	1 091
高黎贡山-碧罗雪山植物多样性调查	
Plant Diversity Survey of Gaoligong - Biluo Snow Mountain	852
李攀的采集记录	
Collection Record of Li Pan	797
深圳植被空间分布	
Spatial Distribution of Plants in Shenzhen	599
专类园引种	
Introduction of Specialized Garden	481
肯尼亚植物志-2019.1 中肯联合科考	
Flora of Kenya - Sino Kenya Joint Scientific Survery	30

chinaXiv:202108.00051v1

chinaXiv:202108.00051v1

乌蒙山植物植被监测-凭证标本采集	
Vegetation Monitoring in Wumeng Mountains - Voucher Specimen Collection	285
西峰药用植物普查	
General Survey of Medicinal Plant in Xifeng Mountain	234
千岛湖植物学野外实习	
Field Practice of Botany in Qiandao Lake	911
植物标本采集与鉴定培训班	
Plant Specimen Collection and Identification Training Class	156
胡光万栽培植物采集	
Collection of Cultivated Plants by Hu Guangwan	246
马达加斯加考察	
Scientific Expedition to Madagascar	176
勐康口岸外来入侵植物调查	
Investigation on Invasive Plants at Mengkang Port	167
浦江县野菜野果资源调查	
Investigation on Wild Vegetable and Fruit Resources in Pujiang County	165

2.2 植物标本的数字化

我们通过对 Kingdonia 系统的日志文件进行分析，获得了图 5 所示的昆明植物研究所标本馆（KUN）两位工人的标本转录效率箱体图。通过箱体图，可以发现相对于人工转录，调用 Biotracks 数据转录标本显示出了以下三个优势。

2.2.1 标本转录的效率明显更高　A、B 工人转录 I 类样本的平均时长分别为 336 和 389 s，A、B 工人转录 II 类样本的平均时长分别为 117 和 124 s。A、B 工人在 II 类样本中，标本的平均转录时长都明显低于 I 类标本，两个工人都有近三倍的转录效率提升。

2.2.2 工人的表现更加稳定　A、B 工人转录 I 类样本的上下限分布范围明显更大，四分位距也更长，中位数与平均值的偏差相对更大，说明工人在转录 I 类样本时，表现并不稳定，有些标本需要耗费较长的时间才能完成转录，有些则可以很快地完成转录；而 A、B 工人在转录 II 类样本时，其转录时长的波动范围要显著小于 I 类样本。这个结果其实可以预见。人工转录标本所需要的时长，往往是和标本标签上的内容多少、字迹辨识的难易程度直接相关，因此单份标本的转录时长会有很大的差异，而调用 Biotracks 数据转录标本，绝大部分的标签内容并不需要人工转录，因此受标签内容性质的影响较小，个人转录效率的波动自然更小。

2.2.3 不同工人之间的转录效率差异更小　A、B 工人转录 I 类样本的平均时长为 336 和 389 s，时差为 53 s；A、B 工人转录 II 类样本的平均时长分别为 117 和 124 s，时差为 7 s；显然在处理 II 类样本时，A、B 两位工人之间的耗时差异不如 I 类样本显著。究其原因，可能是因为对于 II 类样本，工人大多数时间只是在核对所调用的数据与标签内容是否一致，而不用对照标本标签亲自转录文本，对于不同的标签内容，单纯的核对文本内容可能要比手工转录相应的内容更为简单，由此使得不同工人之间的转录效率差异会更小。

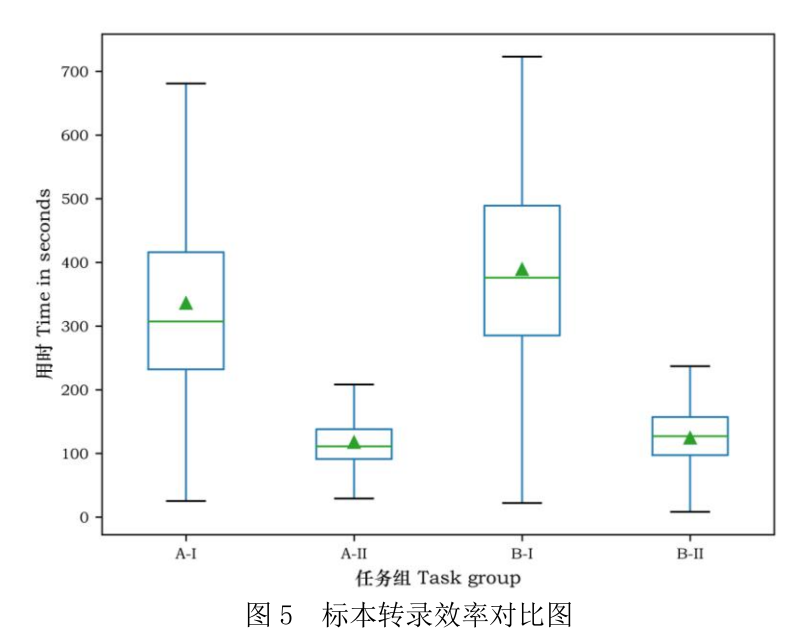


图 5　标本转录效率对比图

Fig.5 Comparison of specimen label transcription efficiency

2.3 标本信息维度的扩展

Biotracks 可以拍摄标本的原生照片并将其与采集信息直接绑定，数据同步至云端之后，照片的网络地址会被写入采集信息的 associatedMedia 字段内，类似 Kingdonia 这样的数字标本馆系统在调取 Biotracks 的采集记录时，就可以将这些照片一并拉取并与相应的数字标本形成关联，从而使得传统标本原本很难呈现的颜色、行为、立体结构、环境等原生信息，现在都可以通过数字标本被展现在研究者面前（图 6）。

chinaXiv:202108.00051v1

ChinaXiv合作期刊

chinaXiv:202108.00051v1

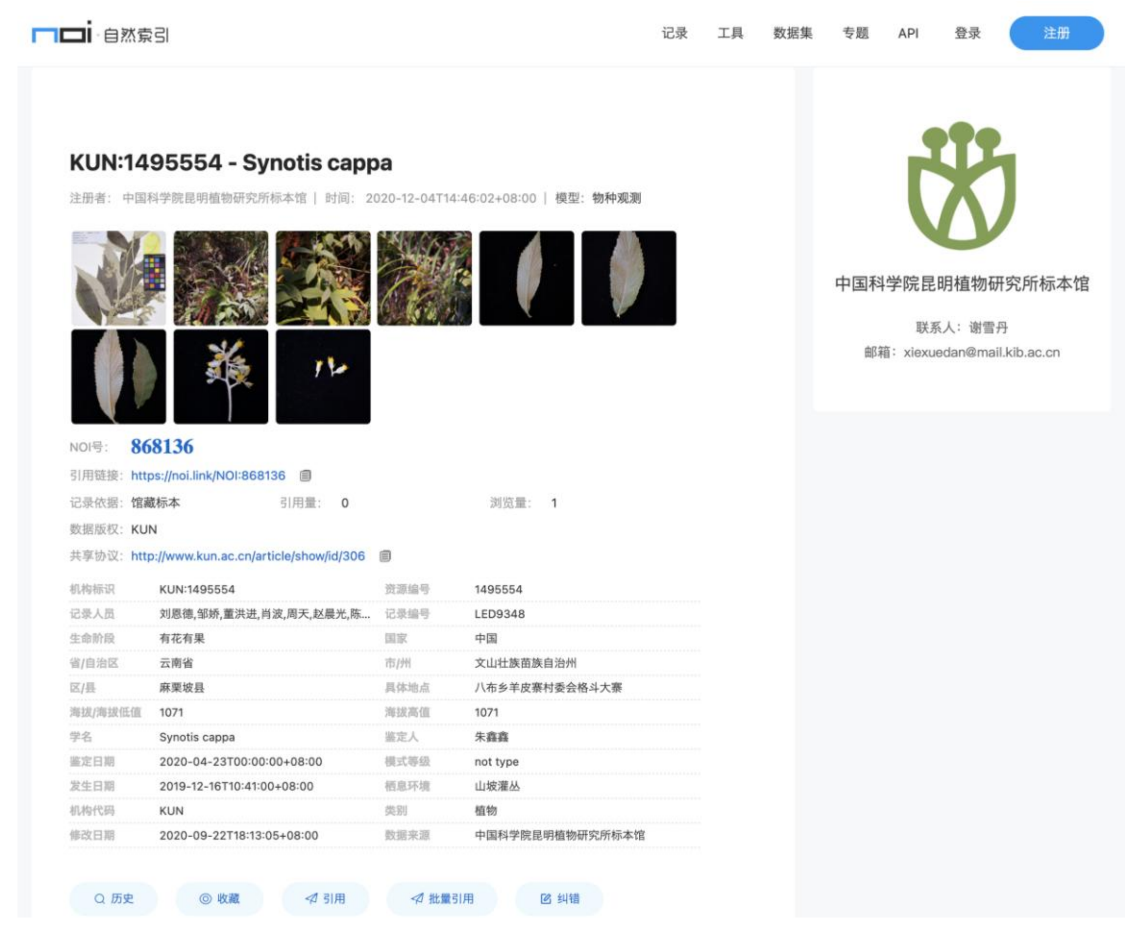


图 6 noi.link 网站上的 KUN:1495554 号数字标本，https://noi.link/NOI:868136

Fig.6 The digital specimen of KUN:1495554, https://noi.link/NOI:868136

实现上述过程，采集者和标本馆并不需要为此增加额外的工作量。但在传统的标本采集流程中，野外照片通常会被每个采集队员保存在各自的电脑或移动硬盘内，如果需要将这些照片和最终的数字标本形成关联，就需要耗费大量的人力和时间用于照片的梳理、编号、汇总和匹配。显然这些工作并不是每个采集者都可以做到，更多的时候这些照片也不会被提交至标本馆，由此可见在原有的工作流程下，希望依靠人力去完成数据的关联本身就是非常困难的。而 Biotracks 确实为此提供了一种实用自然的方案，可以有效解决这一问题。

3 讨论与结论

3.1 Biotracks 为何可以获得广泛应用？

针对标本的采集，在 Biotracks 正式发布前后其实也有一些类似的开放应用出现，比如 collNotes、ColectoR 等（Maya-Lastra, 2016; Powell et al., 2019）。然而这些应用却没有像 Biotracks 这样被广泛地使用。究其原因，手机作为一种新的记录载体，其实是有一些先天劣势：（1）屏幕狭小使得交互空间过于局促，进而会对浏览和编辑信息的体验和效率造成不良影响；（2）手机体型更小，虽然方便携带，但也更容易损坏或丢失，从而影响数据安全；（3）智能手机续航有限，自身电量通常难以维持长时间的户外使用；（4）相较于纸和笔，智能手机的操作仍然有一定的门槛，因此并不是所有人都会接受或适应这一记录方式。

这些缺点中的（3）和（4）可以通过携带移动电源和加强用户培训予以改善，因此最

终真正影响用户是否选择手机记录的主要因素还是在于（1）（2）所述的便捷性和安全性。而这两点恰好是软件优化能够予以充分解决的。通过自动写入、信息复用、交互设计、AI 辅助、字段最小化等多种措施，Biotracks 有效保障了手机 APP 记录的便捷性。而借助网络对数据进行及时同步则可以最大限度地保障数据的安全性。

在此基础上，Biotracks 还有纸质记录或电子表格难以弥补的优势，这包括：（1）使用 Biotracks 不仅可以直接生成电子数据，还可以不依赖于网络和其他外置设备，写入准确的拉丁名以及非常精确的地理坐标、海拔高度、观测时间等信息；（2）智能手机能够直接使用移动运营商提供的 3G、4G、5G 等网络服务，因此可以最大限度地利用互联网增强用户在户外环境中的信息获取能力，比如 Biotracks 的地址解析、AI 鉴定、在线地图、项目热力图都属于网络服务，这些服务不仅提高了记录的完整性，还增强了用户对外界信息的实时感知能力，进而能够提升野外采集工作的品质和效率；（3）手机非常便携，用户几乎随身携带，这使得互联网能够将每个用户更紧密地连接起来，实现更加精细的分工协作；（4）最小化的输入和标准化的输出，可以最大程度地保证数据的规范性和一致性，由 Biotracks 生成的数据表，字段歧义性更小，值和结构更加规范，可以一次生成，多处重用，因此能够大幅降低数据整合和利用的难度。

综上，Biotracks 不仅克服了手机记录的缺点，还进一步发挥出了手机这一载体的优势。使用 Biotracks 进行植物标本的采集，既能够便捷、准确、完整地记录信息，也可以保证数据的规范、一致和安全，同时还实现了更加精细、动态、系统的分工协作，最终使得标本的采集工作更为省时、省力、省心。

3.2 Biotracks 为植物标本的收集建立了一种新的工作流程

除了满足一线工作者信息记录的需求，Biotracks 还能够深度参与从采集任务规划、标本采集到数据的整合、感知、协作、管理，乃至最终标本的数字化，这一整条标本收集流程中各个环节的的优化和重塑（图 7）。在规划采集时，用户可以根据实际需要，建立 Biotracks 虚拟项目，不同采集可以被创建为不同的项目，同一个成员也可以加入不同的项目。在采集标本时，Biotracks 可以记录采集信息，并把相应的数据和文件同步至云端。之后，每个成员记录的数据会被自动整合到相应的项目内，成员可以在项目内分工合作，项目统计功能会实时统计每个成员的贡献，项目地图则可以将整合的数据绘制为热力图，这些功能能够将任务的最新进展传达给成员，成员则可以借助它更好地规划下一步的工作。而在数据管理方面，Biotracks 支持为不同成员设置不同的数据管理权限，整合的数据可以被批量下载后提交给标本馆，标本馆也可以直接通过 Biotracks 的授权接口主动调取采集信息，然后应用于相应的标本数字化。

chinaXiv:202108.00051v1

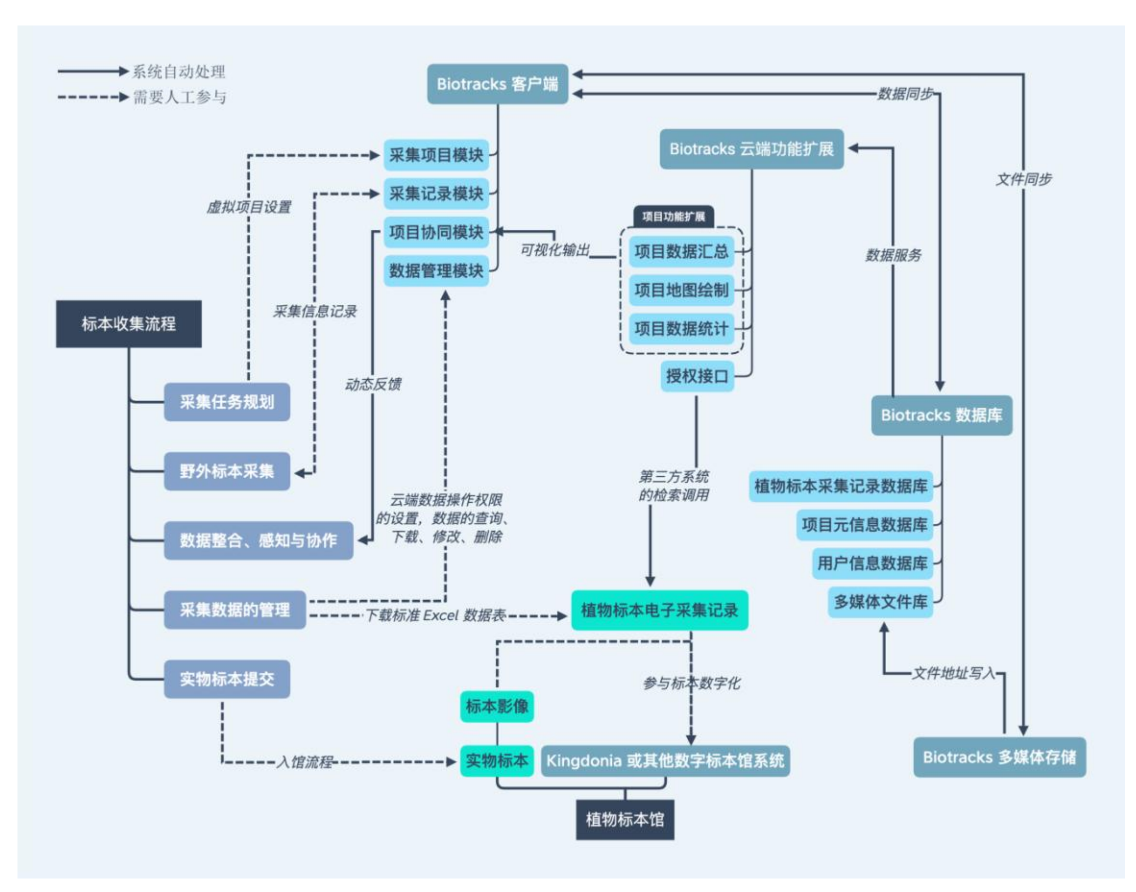


图 7 利用 Biotracks 构建的标本收集流程

Fig.7 Specimen collection process constructed by Biotracks

通过信息化的手段和体系化的设计，Biotracks 不仅可以自然的融入到标本收集流程中的各个环节，还很好地解决了一些传统标本收集流程中固有的问题，这包括：（1）数据规范难以被彻底地贯彻（Biotracks 采用了统一的数据模板记录数据，字段名、值类型、值样式都得到了很好地约束）；（2）数据提交和汇总费时费力（Biotracks 会自动整合项目成员的数据，有权限的成员可以批量下载）；（3）野外工作的成果和进展不能被及时地共享和感知（Biotracks 能够将项目数据共享给所有成员，同时会实时统计各个成员的贡献，还会将采集记录绘制成热力地图展现给成员）；（4）数据无法被长期追踪（Biotracks 将数据存于云端，数据的编辑和修改都可以同步更新至各个成员的最新设备之中）；（5）植物标本采集信息的数字化主要依靠人工转录（标本馆可以导入或调取 Biotracks 记录的采集信息，避免人工二次转录）。

相较于传统的标本收集流程通常会按照时序依次组织各个工作环节，Biotracks 所塑造的流程，则可以让各类人员形成跨越时空的分工合作，它不仅能够提升标本收集过程中各个环节以及整个流程的工作效率，还能够从根本上改变传统数据流的整合方式，因此，应该被视为一种新的标本收集流程，作用于新进标本的收集。对于最终的标本，该模式能够将原生的照片与数字标本自然的形成关联，从而将收藏于标本馆的标本与现实场景联系在一起，这种影响目前我们还无法准确的评估，但对科学研究而言，标本数据一直都是生物多样性数据的金标准（Culley，2013），新的标本若能以这种形式进行大量的富集，其潜质必将是令人期待的。

3.3 Biotracks 为国内植物标本馆的发展带来了公众科学

采集者利用 Biotracks 记录标本采集信息，标本馆调用这些信息转录标本，其中的效率提升实际上是因为标签转录工作被分包给了各个标本采集者。这种借助众人力量解决科学数

chinaXiv:202108.00051v1

ChinaXiv合作期刊

据甚至科学研究问题的模式，被统称为公众科学（张健等，2013；金瑛等，2019）。利用公众科学解决标本馆的问题，其实并不罕见。比如标本标签的转录，国外就有一些公共平台专门帮助各类标本馆和博物馆转录标本标签[比如 Note From Nature（Hill et al., 2012）]，一些博物馆还会建立自己的转录平台以支持公众参与博物馆藏品的标签转录(比如史密森数字志愿者转录中心，https://transcription.si.edu），有些组织甚至还会为此举办国际性的标签转录竞赛[比如 WeDigBio（Ellwood et al., 2018）]。然而在国内，标本馆对于公众科学的关注、应用和研究却很不理想。这其实和国内植物标本馆的发展模式有很大的关系：一方面我国的现代植物学如同其他科学门类一样，属于舶来品，因此与西方社会相比，自然博物的社会基础并不好。另一方面，植物标本馆最初的价值就是收纳标本以支持标本的研究，因此相比面向公众的各类博物馆，它与公众的关系更为疏远。这些因素造成了国内植物标本馆的发展一直以来都很难扎根于公众，导致最终只能依附于植物分类学的发展。然而，由于《中国植物志》、*Flora of China* 等全国性志书已经完成，更由于政策导向等诸多因素的影响，当前中国植物分类学的境遇也不太好，而与之紧密相关的植物标本馆自然也面临着严峻的生存和发展问题（马金双，2010）。

但是国内植物标本馆的发展机遇仍然是存在的。这与科学研究的总体发展趋势有很大的关系。比如近年来标本馆的标本数字化就受到了前所未有的重视。数字标本作为实体标本的延伸，起初只是方便了标本的检索和查阅，然而随着数据的富集，数字标本的价值已经不再局限于植物分类学的研究，而延伸至环境变化、保护生物学、群体遗传学与基因组学，甚至公共健康与安全等领域（Ellwood et al., 2018）。对于严重依赖分类学研究的传统标本馆而言，这不仅扩展了标本馆的生存基础，还为标本馆带来了实实在在的经费支持。比如中国数字植物标本馆（CVH）每年都会支持数十家标本馆的标本数字化工作（刘慧圆等，2017），相应的经费可能不算多，但对于极度缺乏经费支持的标本馆，也是雪中送炭。而数字标本之所以有如此广泛的科学价值，其实并非偶然。2009 年，微软针对数据密集型科学提出了第四研究范式（Hey et al., 2009），之后的十多年随着智能手机，3G、4G、5G 蜂窝网络，计算机硬件，物联网，云存储，云计算，人工智能等技术的发展，大规模数据的获取、管理与利用能力已经获得了质的飞跃（Madden, 2012; McAfee et al., 2012; Ge et al., 2018; Oussous et al., 2018）。这使得基于数据密集型的科研第四范式开始真正从理论发展转向实践应用，数据的价值因此被充分的释放和展现。而标本数据的富集和利用，正是这一趋势下的一个实例。

因此，现代标本馆的科学价值不仅在于所收藏的标本，还在于它能带来的数据价值。与标本馆紧密相关的地面生物多样性数据的收集、整理和利用，属于典型的人力密集型活动，而这正是公众科学所擅长的领域。同时，伴随着我国公众科学素养的不断提高（何薇等，2018），以及国家层面对于生态安全与生物多样性的保护的进一步重视（陆军，2019），国内开展类似公众科学活动的条件已经成熟。而 Biotracks 的核心目标之一，就是希望将各类公众的力量，通过信息技术手段与标本馆形成对接，从而解决标本馆在大规模物种数据的收集、鉴定、提取、利用等方面所遭遇的人力和智力资源匮乏等问题。比如对于野外调查活动，标本馆可以通过 Biotracks 发布科考志愿者招募，从而解决野外调查中人力资源不足的现状（http://www.biotracks.cn/activity）。对于标本的转录，调用 Biotracks 数据转录标本，既能提高效率，还可以降低转录的难度，同样也有益于招募公众参与标本的数字化。而作为未来 Biotracks 的工作重心之一，类似 iNaturelist 和 Biotracks 这样的自然观察平台，聚集了大量的专业人员和分类学爱好者，他们对于物种的鉴定和讨论也完全可以推送给与之相关的数字标本。从而使得传统标本低效、低频率的鉴定和讨论得以改善，并最终推动物种的鉴定和发现向着社区驱动的方向发展（Heberling & Isaac, 2018）。

诸如此类的工作，其实还有很大的扩展空间，通过 Biotracks 我们看到了公众科学对于现代标本馆的价值，而在公众科学活动中，标本馆其实也有自己独特的价值：（1）作为一

chinaXiv:202108.00051v1

chinaXiv:202108.00051v1

个专业机构，首先它有普通个体难以比拟的号召力，这对于公众科学的发起和组织是至关重要的；（2）标本馆可以接触到大量的专业人员，容易和他们产生合作，共同开展公众科学活动；（3）标本馆有大量的植物标本，这些标本的潜在价值是惊人的，借助公众科学的手段可以深度挖掘标本的价值，创作具有更高科学价值的数据产品，比如联合公众从标本图片上提取花期、果期等物候数据集（Willis et al., 2017）;（4）除了深挖标本的价值，标本馆自身的专业能力和区位优势也应该是构建数据价值的重要力量，比如我们是否可以考虑将主导和维护特定区域和特定内容的自然观察数据集作为标本馆未来的核心价值之一呢？

3.4 与类似方案的比较

借助类似 Biotracks 这样的公众科学平台所提供的开放应用采集植物标本，然后再将其与标本馆（博物馆）的标本收藏工作相结合，在国外也有案例可循。iNaturalist（https://www.inaturalist.org）是目前国外发展非常迅速的自然观察平台，用户使用 iNaturalist 的 APP 或网站提交物种的发现记录，iNaturalist 会将这些记录分享给用户社区进行交流和鉴定。2018 年 11 月的一篇公开文献（Heberling & Isaac, 2018）分享了 iNaturalist 针对植物标本收集的一些拓展，总体而言与笔者之前对 Biotracks 的简要说明（徐洲锋，2018）以及在本文中进一步描述的模式基本一致。但两者在具体实现上，也有一些差异，这些差异包括以下几个方面。

3.4.1 项目创建　iNaturalist 的标本采集项目必须在网页端创建，Biotracks 的项目则必须在手机端创建；iNaturalist 对创建者会有一定的要求（至少有 50 条被验证的物种观察记录才可以创建项目），Biotracks 暂时没有这方面的限制。

3.4.2 项目公开　iNaturalist 的标本采集项目是对外公开的，任何用户都可以看到和参与这个项目，Biotracks 则将项目分成了公开、团队、私有三种类型，不同类型有不同的开放度：iNaturalist 会将项目记录开放给所有用户浏览和鉴定，Biotracks 的项目记录只会分享给项目成员。

3.4.3 数据标准　iNaturalist 的标本采集项目，数据模版除了核心字段之外，其他字段可由用户自行定义；Biotracks 也支持定制，但并没有把定制的权限交给用户，针对不同类群，Biotracks 采取了平台定制，用户使用的模式，以确保数据的强一致性。

3.4.4 应用重心　iNaturalist APP 更加轻量，一些关键信息的记录，比如地理坐标有赖于网络或关联图片的支持，且 APP 目前尚不支持海拔高度的获取，项目数据的统计分析和可视化展示也更多的集成在网页端；Biotracks 手机端功能更加丰富，能够不依赖网络记录坐标和海拔信息，同时倾向于将更多的功能集成到移动端，而网页端的功能则相对较弱。

3.4.5 资源标识　iNaturalist 希望在自己的平台上为每条记录生成一个永久的 URL 链接，以固定标本的引证；Biotracks 更希望将采集记录注册到一个开放的第三方资源标识符平台（比如 https://noi.link），以统一不同来源物种记录的数字标识符。

这些差异，有些是环境因素造成的（比如 iNaturalist 相比 Biotracks 更加开放，主要是因为西方国家在公众科学领域的发展更超前，特别是与此有关的数据共享实践要比国内更加成熟，而 Biotracks 目前综合考虑了各方面因素，尚无法贸然主动公开所有用户的数据，但 Biotracks 也在积极探索新的方式，以期在符合国情的情况下，尽可能促进数据分享）。有些是由于发展理念不同造成的差异（比如资源的标识，Biotracks 认为领域需要有统一的数字标识符，因此并不赞成平台将自身的固定地址应用在标本资源上）。有些则是由各自所处的发展阶段造成的（Biotracks 先期更加侧重移动端的发展，但未来也需要加强 WEB 端的能力）。但整体而言，这两个平台针对标本的收集所构建的方案可谓不谋而合。当然，相比 iNaturalist，Biotracks 的体量还非常的小，但我们也希望 Biotracks 能够在未来中国植物标本馆的发展历程上留下自己的身影。

参考文献：

ChinaXiv合作期刊

CULLEY TM, 2013. Why vouchers matter in botanical research[J]. Appl Plant Sci, 1(11): 1300076.
DARU BH, PARK DS, PRIMACK RB, et al., 2017. Widespread sampling biases in herbaria revealed from large-scale digitization[J]. New Phytol, 217(2): 939-955.
DRINKWATER RE, CUBEY RWN, HASTON EM, 2014. The use of optical character recognition (OCR) in the digitisation of herbarium specimen labels[J]. PhytoKeys, 38(38): 15-30.
ELLWOOD ER, KIMBERLY P, GURALNICK R, et al., 2018. Worldwide engagement for digitizing biocollections (WeDigBio): the biocollections community’s citizen-science space on the calendar[J]. Bioscience, 68: 112–124.
GE M, BANGUI H, BUHNOVA B, 2018. Big data for internet of things: a survey[J]. Future Gener Comput Syst, 87: 601–614.
GRIES C, GILBERT E, FRANZ N, 2014. Symbiota—A virtual platform for creating voucher-based biodiversity information communities[J]. Biodivers Data J, 2: e1114-24.
HE W, ZHANG C, REN L, et al., 2018. Public attitudes towards and understanding of science and technology in China: Based on the results of vivic scientific literacy survey 2018[J]. Stud Sci Popul, 13(77): 51-60.[何薇，张超，任磊，等，2018. 中国公民的科学素质及对科学技术的态度———2018 年中国公民科学素质抽样调查结果[J]. 科普研究，13(77)：51-60.]
HEBERLING JM, ISAAC BL, 2018. iNaturalist as a tool to expand the research value of museum specimens[J]. Appl Plant Sci, 6(11): e01193.
HE P, CHEN J, CAI L, et al., 2021. Import supporting role of biological specimen in biodiversity conservation and research[J]. Bull Chin Acad Sci, 30(4): 425–435.[贺鹏，陈军，蔡磊，等，2021. 生物样本：生物多样性研究与保护的重要支撑[J]. 中国科学院院刊，30(4)：425–435.]
HEY AJG, TANSLEY S, TOLLE KM, et al., 2009. The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft Res.
HILL A, GURALNICK R, SMITH A, et al., 2012. The notes from nature tool for unlocking biodiversity records from museum records through citizen science[J]. ZooKeys, 209(209): 219-233.
HOBERN D, BAPTISTE B, COPAS K, et al., 2019. Connecting data and expertise: A new alliance for biodiversity knowledge[J]. Biodivers Data J, 7: e33679.
JAMES SA, SOLTIS PS, BELBIN L, et al., 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research[J]. Appl Plant Sci, 6(2): e1024.
JIN Y, ZHANG XL, HU ZH, 2019. Development and challenge of citizen science[J]. Libr Inform Serv, 63(13): 28-33.[金瑛，张晓林，胡智慧，2019. 公众科学的发展与挑战[J]. 图书情报工，63(13)：28-33.]
LIU HY, QIN HN, LI M, 2017. Plant specimen resource sharing platform and plant specimen digital capability construction[J]. E-Sci Technol Appl, 8(4): 13-23.[刘慧圆，覃海宁，李敏，2017. 植物标本资源共享平台与标本数字化能力建设[J]. 科研信息化技术与应用，8(4)：13–23.]
LU J, 2019. Strengthening national ecological security is the need of realizing sustainable development in China[J]. Environ Prot, 47(8): 8-12.[陆军，2019. 强化国家生态安全是我国实现可持续发展的需要[J]. 环境保护, 47(8)：8-12.]
MA JS, 2010. Current status and challenges of Chinese plant taxonomy[J]. Chin Sci Bull, 59(6): 510–521.[马金双，2010. 中国植物分类学的现状与挑战[J]. 科学通报，59(6)：510-521.]

chinaXiv:202108.00051v1

ChinaXiv合作期刊

MADDEN S, 2012. From databases to big data[J]. IEEE Internet Comput, 16(3): 4-6.
MAYA-LASTRA CA, 2016. ColectoR, a digital field notebook for voucher specimen collection for smartphones[J]. Appl Plant Sci, 4(7): 1600035-5.
MCAFEE A, BRYNJOLFSSON E, DAVENPORT TH, et al., 2012. Big data: The management revolution[J]. Harvard Bus Rev, 90(10): 60-68.
MORRISON SA, SILLETT TS, FUNK WC, et al., 2017. Equipping the 22nd-century historical ecologist[J]. Trend Ecol Evolut, 32(8): 578-588.
OUSSOUS A, BENJELLOUN F-Z, LAHCEN AA, et al., 2018. Big data technologies: a survey[J]. J King Saud Univ-Comput Inform Sci, 30(4): 431–448.
POWELL C, MOTLEY J, QIN H, et al., 2019. A born-digital field-to-database solution for collections-based research using collNotes and collBook[J]. Appl Plant Sci, 7(8): e11284.
SCHINDEL DE, COOK JA, 2018. The next generation of natural history collections[J]. Plos Biol, 16(7): e2006125.
SFORZI A, TWEDDLE J, VOGEL J, et al., 2018. Citizen science and the role of natural history museums[M]// HAKLAY M. Citizen science Innovation in open science: Society and Policy: UCL Press: 429-444.
SOLTIS PS, 2017. Digitization of herbaria enables novel research[J]. Amer J Bot, 104(9): 1281-1284.
VISSERS J, BOSCH, DEN FV, et al., 2017. Scientific user requirements for a herbarium data portal[J]. PhytoKeys, 78(78): 37-57.
WANG LS, CHEN B, JI LQ, et al., 2010. Progress in biodiversity informatics[J]. Biodivers Sci, 18(5): 429-443.[王利松，陈彬，纪力强，等，2010. 生物多样性信息学研究进展[J]. 生物多样性，18(5)：429–443.]
WILLIS CG, ELLWOOD ER, PRIMACK RB, et al., 2017. Old plants, new tricks: Phenological research using herbarium specimens[J]. Trend Ecol Evolut, 32(7): 531-546.
WILLIS CG, LAW E, WILLIAMS AC, et al., 2017. CrowdCurio: an online crowdsourcing platform to facilitate climate change studies using herbarium specimens[J]. New Phytol, 215(1): 479-488.
XU ZF, 2018. Kingdonia project: A herbarium based citizen science practice[J]. E-Sci Technol Appl, 8(4): 97-105. [徐洲锋，2018. 结合公民科学的 Kingdonia 系统工作平台的构建与应用[J]. 科研信息化技术与应用，8(4)：97-105.]
ZHANG J, CHEN SB, CHEN B, et al., 2013. Citizen science: Integrating scientific research, ecological conservation and public participation[J]. Biodivers Sci, 21(6): 738-749.[张健，陈圣宾，陈彬，等，2013. 公众科学:整合科学研究、生态保护和公众参与[J]. 生物多样性，21(6)：738–749.]

chinaXiv:202108.00051v1